The relation between stature and long bone length in the Roman Empire

Klein Goldewijk, Geertje; Jacobs, Jan

university of groningen / faculty of economics and business

13002-EEF

The relation between stature and long bone length in the Roman Empire

Geertje Klein Goldewijk
Jan Jacobs

university of groningen / faculty of economics and business

SOM is the research institute of the Faculty of Economics & Business at the University of Groningen. SOM has six programmes:
- Economics, Econometrics and Finance
- Global Economics & Management
- Human Resource Management & Organizational Behaviour
- Innovation & Organization
- Marketing
- Operations Management & Operations Research

Research Institute SOM
Faculty of Economics & Business
University of Groningen

Visiting address:
Nettelbosje 2
9747 AE Groningen
The Netherlands

Postal address:
P.O. Box 800
9700 AV Groningen
The Netherlands

T +31 50 363 7068/3815

www.rug.nl/feb/research

The relation between stature and long bone length in the Roman Empire

Geertje Klein Goldewijk
University of Groningen
geertjekleingoldewijk@gmail.com

Jan Jacobs
University of Groningen

The relation between stature and long bone length in the Roman Empire

Geertje M. Klein Goldewijk, Groningen Institute of Archaeology, University of Groningen
[geertjekleingoldewijk@gmail.com]
Jan P.A.M. Jacobs, Faculty of Economics and Business, University of Groningen

Version February 2013

Abstract

Stature is increasingly popular among economic historians as a proxy for (biological) standard of living. Recently, researchers have started branching out from written sources to the study of stature from skeletal remains. Current methods for the reconstruction of stature from the skeleton implicitly assume fixed body proportions. We have tested these assumptions for a database containing over 10,000 individuals from the Roman Empire. As it turns out, they are false: the ratio of the length of the thigh bone to the length of the other long bones is significantly different from those implied in the most popular stature reconstruction methods. Therefore, we recommend deriving a proxy for living standards from long bone length instead of reconstructed stature.

Key words: body proportions, living standards, long bones, Roman Empire, stature.

Acknowledgements: This research has been funded by NWO, Toptalent grant nr. 021.001.088. We thank Wim Jongman, Gerard Kuper, and Vincent Tassenaar for their help and comments.

1. Introduction

Stature is increasingly popular among economic historians as a proxy for (biological) standard of living (Steckel 2009). The better a child is fed, the taller it can grow. That not only depends upon how much it eats, but also on how much it needs: the harder a child has to work, the more fuel its muscles need; the more pathogens it encounters, the more of an effort it takes to ward them off; the more poorly it is housed and clad, the more energy it has to spend to keep warm. If a child is short on nutrients, it has to cut on growth. Its low nutritional status is reflected in a small stature. On the level of the individual, genes play an important role, but on a group level the genetic influences cancel each other out. Average stature thus is related to the quality and quantity of food, clothing, housing, disease and work load. That makes it a good proxy for overall living standards.

In economic history, the vast majority of stature research is based on written sources on height, such as conscription lists. However, written data is only available for more recent periods. Data from human skeletal remains can supplement the written sources. Koepke and Baten (2005) study the development of living standards in Europe from the first to the eighteenth century CE using stature from skeletons. Steckel collects several skeletal indicators of health, including stature, in an effort to elucidate the development of living standards in Europe and the America's in the last ten thousand years (see Steckel and Rose, 2002 for some of the first results). Koca Özer et al. (2011) and De Beer (2004) use skeletal evidence to study the secular change in height in Turkey and the Netherlands, respectively.

For our research into living standards in the Roman Empire, we collected published and unpublished osteological reports on human skeletal remains found in the Roman Empire, and dated between 500 BCE and 750 CE. Stature reconstruction is a standard part of osteological analysis, and most skeletal reports contain some stature figures. These figures, however, have been produced using a wide array of stature reconstruction methods, and they cannot be lumped together just like that.

In this article, we will test the ten most popular methods for the reconstruction of stature from the skeleton. We will calculate the long bone length proportions implied by these methods, and test these against the long bone lengths proportions in Roman period skeletons. As a result, we will propose an alternative approach: we advise not to attempt the reconstruction of stature, but to study the development of long bone length instead.

The remainder of this article is structured as follows. Section 2 discusses the extant stature reconstruction methods. Section 3 introduces our database, and the type of analysis that we use. Section 4 presents our results, the implications of which are discussed in section 5. Section 6 contains a short conclusion.

2. Reconstruction of stature from the skeleton

Most skeletons that are found cannot be measured from head to heel. They are incomplete, or the bones are out of position. Fortunately, stature can be reconstructed from the long bones, the large bones of the limbs. In the nineteenth century, scientists already assumed that there is a relation between the length of the body and that of the limbs. Rollet (1888) measured 100 dissecting room cadavers from Lyon, and calculated the average length of each long bone in men and women of a similar stature. Pearson (1899) performed regression analyses on Rollet´s data, and came up with two sets of stature reconstruction formulae, one for men and one for women, which can be used to calculate stature from the length of a single long bone (see table 1).

Pearson´s work set the standard for twentieth century studies into the relation between long bone length and stature. All perform regression analyses, albeit on data from different populations: Breitinger (1937) measured male students and athletes living in Germany in the 1920´s; Bach (1965) provided the matching formulae for females from women living in Jena in the 1960´s; Eliakis et al. (1966) studied university dissecting room cadavers from Athens, Telkkä (1950) studied those from Helsinki; Olivier wrote a series of articles on western Europeans deported in the Second World War (Olivier, 1963; Olivier and Tissier, 1975; Olivier et al., 1978); Dupertuis and Hadden (1951) published different sets of formulae for whites and blacks, based on an early twentieth century collection of skeletons from Ohio; Trotter and Gleser (1952, 1958) complemented that dataset with American soldiers killed in the Pacific during the Second World War and the Korean War.

All these regression studies come up with different sets of formulae. And the choice of formula has a significant effect on the resulting stature figure. For example, the average length of the male thigh bone or femur in our database is 450 millimeter. This yields a predicted stature between 165.3 cm (Trotter and Gleser, 1952, for blacks) and 172.8 cm (Eliakis et al., 1966). In part, this is due to differences in measurement methodology: some measure the bones when they are ´fresh´, others wait for them to dry; some take maximum bone length, others prefer the length to be measured in the anatomical position; some researchers have stature measurements taken during life, others have to make do with cadavers lying on a table or suspended from the ceiling. However, when this diversity is accounted for, the discrepancy remains more than 5 centimeters.

Physical anthropologists soon remarked upon these differences in body proportions. They ascribed it to genes, and they devised separate sets of formulae for different peoples (‘races’). More recently, they realized that even when the genetic composition of a population stays more or less the same, body proportions can still change. The formulae that Trotter and Gleser published on Second World War victims (Trotter and Gleser, 1952) proved not to be valid anymore for those killed during the Korean War, six to ten years later (Trotter and Gleser, 1958). ´Stature and its relationship to long bone length are in a state of flux´, Trotter and Gleser (1958, p. 122) conclude, and ´equations for estimation of stature should be derived anew at opportune intervals.´ Apparently, body proportions do not only depend upon genes, but also on the environment. Stature reconstruction formulae can therefore only be applied to the population for which they were calculated, or one that is very similar in its genetic composition and its way of life.

As all stature reconstruction methods are based upon late nineteenth or even twentieth century populations, it is hard to pick a method for a population from before that period. In the past, physical anthropologists working with archaeological samples simply followed national tradition: the Germans used the formulae by Breitinger (1937) and Bach (1965); the French employed the tables of Manouvrier (1892, 1893) (based on a subset of the Rollet (1888) data); the Americans turned to the publications of

Trotter and Gleser (1952,1958). Nowadays, more and more physical anthropologists find this praxis unsatisfactory. They emphasize that the stature figures they provide are nothing but a rough approximation of actual body size. They deplore the lack of comparability of estimates made with different methods, and they apply various sets of formulae side-by-side (e.g. Becker, 1999; Lazer, 2009; Rühli et al., 2010). As 'present-day formulae may introduce a systematic bias in estimates of stature of individuals of past generations' (Trotter and Gleser, 1958, p. 116), we must make sure to use the right set of formulae for the Roman period.

3. Material and method

For our study of living standards in the Roman Empire, we collected published and unpublished osteological reports on human skeletal remains found in the Roman Empire, and dated between 500 BCE and 750 CE (Klein Goldewijk, forthcoming). The Roman stature database contains over 10,000 adult men and women born between 500 BCE and 750 CE and buried in the territory of the Roman Empire at its largest extent. It includes all prevailing length measures of all six long bones, over 35,000 in total (see table 2).

We do not know the stature of the men and women in our database. We only know the length of one or more of their long bones. Therefore, we have no way to find out which method renders the correct body heights. We can only search for a method that provides us with a proxy that is internally consistent: that always provides us with the same stature figure, regardless of the long bone that the estimate is based upon.

we need a stature reconstruction method that fits the body proportions of the skeletons in our Roman sample population.

As the femur is the most numerous long bone, we have made it the yardstick against which the other bones are judged. We estimate the relation between femur length and the length of the other five long bones in our database, and we compare that to the long bone length proportions predicted by the extant stature reconstruction methods.

Let us explain that in more detail with the Pearson (1899) formulae that we introduced above. Pearson found the following relation between male stature and femur length: *stature* = 81.306 + 1.880 * *femur*. He also found an association between male stature and humerus length: *stature* = 70.641 + 2.894 * *humerus*. In both formulae the part before the equals sign is the same (*stature*). Therefore, we can equate the two formulae to each other: 81.306 + 1.880 * *femur* = 70.641 + 2.894 * *humerus*. This boils down to: *femur* = *–5.673* + 1.539 * *humerus*, which we can compare to the ratio of femur to humerus length in our database.

We estimate the long bone length proportions in the Roman stature database using a standard (OLS) linear regression analysis. We run the regressions for men and women independently, as most stature reconstruction methods have separate sets of formulae for men and women, and as there are important biological reasons to suspect that body proportions vary by sex. We assume that the relation between the lengths of two bones is linear, in line with the stature reconstruction methods that we are testing. Hence, we choose to ignore the fact that a few of the estimated models fail to pass the Ramsey RESET test, suggesting that a quadratic or an exponential model might have a better fit (see tables 3, last column). We tested for heteroskedasticity using White's heteroskedasticity test (see tables 3, penultimate column). If homoskedasticity is rejected, we adjust the standard deviations accordingly. We calculate the 95% confidence interval for each parameter, and compare the resulting values with those from the stature reconstruction formulae.[1] When both the constant and the slope parameter from a stature reconstruction method fall within the 95% confidence interval from our database, we test both parameters together using the Wald test.

We share some of the worries expressed by Sjøvold (1990) about the use of OLS regression in stature reconstruction research. However, we feel that his alternative, Reduced Major Axis analysis, does not solve the endogeneity problem. Instead, we have done a much more extreme robustness check: we ran all regressions described in this article 'the other way round', i.e. with the femur on the right side of the equation.

We test the ten stature reconstruction methods that are most popular among physical anthropologists studying Roman period skeletons. We restrict ourselves to the formulae for ´whites´, as the inhabitants of the Roman Empire, however genetically diverse, can for the large majority be expected to be ´Caucasian´. We make an exception for Trotter and Gleser´s formulae for blacks, as they perform well in previous studies into stature reconstruction in Roman period skeletons (Becker, 1999; Giannecchini and Moggi-Cecchi, 2008). We also include the formulae for blacks by Dupertuis and Hadden (1951), as their sample population overlaps with the one used by Trotter and Gleser (1952).

[1] The 95% confidence intervals of the constant and slope parameters of the extant stature reconstruction methods cannot be computed, because the relevant statistics have not been published.

4. Results

The results of the linear regression analyses are reported in table 3a and 3b. For example, for the men in our database, the relation between femur and humerus length turned out to be:

$$(1) \qquad \textit{femur} = \underset{(7.005)}{73.239} + \underset{(0.022)}{1.164} * \textit{humerus}$$

n = 1398 R^2 = .683

White heteroskedasticity: p = .038

Under the parameters, between parentheses, is the standard error of the estimate. As homoskedasticity is rejected at the 5% level (White: p = .038), we use robust White-adjusted standard errors, which usually are somewhat larger than the regular ones. These standard errors are used to compute the confidence interval for each of the parameters. As the number of observations is large enough to assume normality, we multiply them with 1.96 to arrive at the 95% confidence interval (see table 4a):

$$(2) \qquad \textit{femur} = 59.509 \text{ to } 86.969 + 1.121 \text{ to } 1.207 * \textit{humerus}$$

Recall that the predicted ratio of femur to humerus length implicit in Pearson's set of formulae for males is:

(3) $femur = -5.673 + 1.539 * humerus$

Both the constant and the slope parameter fall outside the confidence intervals of equation (2). Thus, the Roman men in our database do not fit Pearson's (1988) stature reconstruction formulae for femur and humerus.

This way, we have tested all ten stature reconstruction formulae, for all bone measurements. The results can be found in table 4. The upper and lower boundaries of the 95% confidence intervals are in the first and last columns of tables 4. The middle columns contain the values derived from the stature reconstruction formulae. Those that fall within the confidence interval are printed in bold type. For the men (table 4.a), they do so only occasionally; for the women (table 4.b), they are more often correct. When both the constant and the slope parameter from a stature reconstruction method fall within the 95% confidence interval from our database, we tested both parameters together using the Wald test. In all cases, the parameter values were significantly different from those for the Roman stature database (p = .000). Thus, not a single stature reconstruction method fits the Roman bone length data.

The results of our robustness check (see section 3) are similar: the body proportions implicit in the stature reconstruction formulae do not fit those in the Roman stature database (see table 5 and 6). There are two exceptions: the ratio between male femur length nr. 2 and tibia length nr. 1b as predicted by Pearson, and the ratio between female femur length nr. 2 and tibia length nr. 1a, also by Pearson. However, as all other long bone length proportions do not match, Pearson still does not make a suitable stature reconstruction method.

5. Discussion

Several physical anthropologists have tried to determine which stature reconstruction method serves best for a particular skeletal population. Two studies concern the Roman period. Becker (1999) measured long bone length and body length *in situ* in fifth to third century BCE graves in Satricum, Italy. He concludes that Trotter and Gleser's (1952) formulae for blacks are best. Unfortunately, only twenty of the 179 burials were well enough preserved to allow measurements being taken.[2] Preservation was too poor for regular sex determination, so that Becker had to rely on odontometrics and bone robusticity. While Becker must be commended for working with such problematic material, we fear that the small sample size, the difficulties in taking some of the measurements, and the uncertainty of some of the sex assessments weaken his argument. Besides, as Becker is well aware of, his study pertains to a single cemetery, so its validity is quite limited.

The second study has a wider geographical and temporal scope. Giannecchini and Moggi-Cecchi (2008) sexed and measured over one thousand Iron Age, Roman and Medieval skeletons from central Italy. They selected all skeletons with at least one femur, tibia, humerus and radius, and then for each individual calculated stature four times, i.e., from each bone separately. The closer the four stature estimates are to each other, the better they believe the stature reconstruction method to be. They recommend using Pearson (1899), or Trotter and Gleser´s (1952) formulae for blacks. Unfortunately, the sample sizes of Giannecchini and Moggi-Cecchi are fairly small. Only 179 male and 132 female skeletons still have the four long bones required to qualify for the test, which

[2] Becker (1999) himself writes that his sample size is twenty four, but in four skeletons body length has been measured from field drawings made by archaeology students (Becker (1999), p. 237, table 1), which cannot be too reliable.

seems a bit meager for a time span of almost 2,500 years. The sample size for the Roman period (defined by them as 500 BC to 500 BCE), is 50 men and 38 women only. Second, Giannecchini and Moggi-Cecchi only provide a ranking of stature reconstruction methods, not an absolute judgment: they say which method performs best, but they do not say if the best is also good enough.

We have tested the ten most popular stature reconstruction methods for a database of over 10,000 skeletons from all over the Roman Empire. The results are unequivocal: the long bone length proportions in the Roman stature database do not fit those implicit in the stature reconstruction formulae. Therefore, we feel it is best not to try and reconstruct Roman body length at all, and stick to the information that we have and that we can rely on: the raw data, the long bone lengths.

We suspect similar problems with the reconstruction of stature in other pre-modern skeletal populations. Stature reconstruction formulae are specific for a certain time and place. They should only be applied to the population they were calibrated for, or one that is much alike. It will not do to support the choice for a set of reconstruction formulae for skeletons from the first to the eighteenth century CE with a study pertaining to the Stone Age, as Koepke and Baten (2005) do, referring to Formicola (1993). If the stature reconstruction method does not fit the population that it is used upon, the resulting figures may be off, seriously affecting conclusions about height.

Long bone length is not only a more reliable indicator of living standards than reconstructed stature, it may be a more sensitive one as well. In times of need the development of the trunk, containing most vital organs, may be privileged over that of the limbs. Living conditions may therefore have a stronger effect on long bone length than on body length. Indeed, in the vast majority of stature reconstruction formulae the slope parameter is larger than one, suggesting that within a single population, long bone length varies more than stature does. In the one case where we can compare a single population diachronically, the studies of Trotter and Gleser on American soldiers killed in the Second World War (Trotter and Gleser, 1952) and in the Korean War (Trotter and Gleser, 1958), average stature increases, but the majority of the slope parameters decreases over time (see also Trotter and Gleser, 1958, figure 1 p. 94 and figure 2 p. 96). This suggests that long bone length has gone up more than total body length, and that long bone length is a more sensitive indicator of the change in living standards.

Bone length is harder to collect than reconstructed stature, as the raw data often is not included in the published reports, and physical anthropologists sometimes are reluctant to share their hard-earned data, or the original records have long been lost. Still, a smaller, good-quality database is to be preferred to a larger one filled with erroneous information. What we lose in sample size, we gain in the reliability of our data.

6. Conclusion

Stature normally cannot be measured from the skeleton in the grave. It must be reconstructed from the length of the long bones, but the methods with which that can be done are specific for a certain time and place. The most popular stature reconstruction methods are based on (early-)modern populations. This paper has shown that existing stature reconstruction methods do not fit one particular pre-modern population, that of the Roman Empire. We therefore recommend using long bone length rather than reconstructed stature as (a base for) an indicator of living standards.

Bibliography

Bach, H., 1965. Zur Berechnung der Körperhohe aus den langen Gliedmassenknochen weiblicher Skelette. *Anthropologischer Anzeiger* 29, 12-21.

Becker, M.J., 1999. Calculating stature from *in situ* measurements of skeletons and from long bone lengths: an historical perspective leading to a test of Formicola´s hypothesis at 5th century BCE Satricum, Lazio, Italy. Rivista di Antropologia (Roma) 77, 225-247.

de Beer, H., 2004. Observations on the history of Dutch physical stature from the late-Middle Ages to the present. Economics and Human Biology 2, 45-55.

Breitinger, E., 1937. Zur Berechnung der Körperhöhe aus den langen Gliedmassenknochen. Anthropologischer Anzeiger 14, 249-274.

Dupertuis, C. W., Hadden, J.A., 1951. On the reconstruction of stature from the long bones. American Journal of Physical Anthropology 9, 15-54.

Eliakis, C., Eliakis, C.E., Iordanidis, P., 1966. Sur la determination de la taille d'après la mensuration des os longs. Annales de Médicine Legale 46, 403-421.

Formicola, V., 1993. Stature reconstruction from long bones in ancient population samples: An approach to the problem of its reliability. American Journal of Physical Anthropology 90, 351-358.

Fully, G., Pineau, H., 1960. Détermination de la stature au moyen du squelette. Annales de Médecine Légale 40, 145-154.

Giannecchini, M., Moggi-Cecchi, J., 2008. Stature in archaeological samples from Central Italy: methodological issues and diachronic changes. American Journal of Physical Anthropology 135, 284-292.

Jantz, R.L., Hunt, D.R., Meadows, L., 1995. The measure and mismeasure of the tibia: implications for stature estimation', Journal of Forensic Sciences 40, 758-761.

Klein Goldewijk, G.M., forthcoming. Stature and the standard of living in the Roman Empire, PhD thesis, University of Groningen.

Koca Özer, B., Sağir, M., Özer, İ., 2011. Secular changes in the height of the inhabitants of Anatolia (Turkey) from the 10th millennium B.C. to the 20th century A.D.. Economics and Human Biology 9, 211-219.

Koepke, N., Baten, J., 2005. The biological standard of living in Europe during the last two millennia. European Review of Economic History 9, 61-95.

Lazer, E., 2009. Resurrecting Pompeii. London: Routledge.

Manouvrier, L., 1892. Détermination de la taille d´après les grand os des membres. Revue Mensuelle de l´École d´Anthropologie de Paris 2, 227-233.

Manouvrier, L., 1893. La détermination de la taille d´après les grand os des membres. Mémoires de la Société d´Anthropologie de Paris, 2nd series, 4, 347-402.

Martin, R., 1928. Lehrbuch der Anthropologie in systematischer Darstellung, mit besonderer Berücksichtigung der anthropologischen Methoden. Jena: Gustav Fischer.

Olivier, G., Aaron,C., Fully, G., Tissier, G., 1978. New estimations of stature and cranial capacity in modern man. Journal of Human Evolution 7, 513-518.

Pearson, K., 1899. Mathematical contributions to the theory of evolution: V. On the reconstruction of stature of prehistoric races. Philosophical Transactions of the Royal Society in London, series A, 192, 169-244.

Rollet, E., 1888. De la mensuration des os longs des membres. Lyon: Storck.

Selinsky, P., 2004. An osteological analysis of human skeletal material from Gordion, Turkey, MA thesis at the Department of Anthropology, University of Pennsylvania.

Sjøvold, T., 1990. Estimation of stature from long bones utilizing the line of organic correlation. Human Evolution 5, 431–447.

Steckel, R.H., 2009. Heights and human welfare: recent developments and new directions. Explorations in Economic History 45, 1-23.

Steckel, R.H., Richard, H., Rose, J.C., 2002. The backbone of history: health and nutrition in the Western Hemisphere. Cambridge: Cambridge University Press.

Telkkä, A., 1950. On the prediction of human stature from the long bones. Acta Anatomica 9, 103-117.

Trotter, M., Gleser, G.C., 1952. Estimation of stature from long bones of American whites and negroes. American Journal of Physical Anthropology 10, 463-514.

Trotter, M., Gleser, G.C., 1958. A re-evaluation of estimation of stature based on measurements taken during life and long bones after death. American Journal of Physical Anthropology 16, 79-124.

Table 1 An example of a stature reconstruction method: the formulae by Pearson (1899)

men (n = 50)	women (n = 50)
stature = 81.306 +1.880 * *femur*	*stature* = 72.844 + 1.945 * *femur*
stature = 78.664 + 2.376 * *tibia*	*stature* = 74.774 + 2.352 * *tibia*
stature = 70.641 + 2.894 * *humerus*	*stature* = 71.475 + 2.754 * *humerus*
stature = 85.925 + 3.271 * *radius*	*stature* = 81.224 + 3.343 * *radius*

Notes:

a. Formulae for the reconstruction of living stature from dry bones, Pearson (1899), 196.
b. All bone measures are nr. 1 measurements as specified by Martin (1928).

Table 2 Number of observations in the Roman stature database

			men	women
number of individuals		minimum[a]	5745	4261
		maximum	7879	5926
leg bones	femur	measure nr. 1[b]	4198	3164
		measure nr. 2	1789	1306
	tibia	measure nr. 1	3522	2537
		measure nr. 1a	219	74
		measure nr. 1b	738	585
	fibula	measure nr. 1	746	546
arm bones	humerus	measure nr. 1	3564	2554
		measure nr. 2	715	485
	radius	measure nr. 1	2922	2121
		measure nr. 1b	228	159
		measure nr. 2	337	227
	ulna	measure nr. 1	1928	1316
		measure nr. 2	304	225
sum of bone measures			21283	15339

Notes:

a. We do not know how many individuals the database contains exactly, as some publications only mention the average long bone length of a group of skeletons. If we find an average value for, say, four female left femora and another average value for three female left humeri, we do not know whether these three humeri belong to women who also had a femur to be measured, or if they are three different women entirely. Unless the physical anthropologists mention the number of individuals separately, sample size could be anywere between four and seven.

b. Bone measure numbers refer to Martin (1928).

Table 3a Results of linear regression analysis on bones in Roman stature database (men)

	constant		slope		model			
bones[a]	estimate[b]	S.E.[d]	estimate[b]	S.E.[d]	n	adj. R^2	hetero-skedasticity[c]	Ramsey RESET test[e]
fem1 and tib1[d]	117.827	6.220	.913	.017	1349	.737	.000	.794
fem1 and tib1a	67.477	19.443	1.036	.053	96	.801	.219	.649
fem1 and tib1b	111.806	9.764	.935	.027	432	.739	.225	.160
fem1 and fib1	118.491	14.228	.931	.040	343	.698	.036	.791
fem1 and hum1	73.239	7.005	1.164	.022	1398	.683	.038	.875
fem1 and hum2	59.887	12.709	1.226	.040	571	.681	.007	.224
fem1 and rad1	122.190	8.183	1.341	.033	1127	.633	.000	.087
fem1 and rad1b	66.256	21.097	1.592	.085	153	.695	.529	.613
fem1 and rad2	88.629	19.855	1.573	.084	171	.670	.778	.934
fem1 and uln1	105.358	10.187	1.302	.038	762	.606	.588	.057
fem1 and uln2	136.904	24.022	1.347	.100	160	.529	.003	.014
fem2 and tib1	110.757	8.374	.923	.023	751	.733	.041	.327
fem2 and tib1a	80.733	18.229	.991	.049	112	.783	.588	.559
fem2 and tib1b	114.688	10.699	.916	.029	375	.723	.121	.253
fem2 and fib1	123.409	17.208	.908	.047	223	.622	.166	.908
fem2 and hum1	76.698	9.496	1.148	.029	727	.682	.172	.102
fem2 and rad1	122.212	9.534	1.336	.039	607	.663	.669	.503
fem2 and rad1b	98.432	27.043	1.444	.109	80	.687	.989	.824
fem2 and uln1	102.267	12.795	1.304	.048	476	.613	.417	.621

Notes:

a) Bone measures are abbreviated in the following way: fem = femur, tib = tibia, fib = fibula, hum = humerus, rad = radius, uln = ulna; Numbers refer to bone measure numbers in Martin (1928).

b) All parameter estimates are significant at the 1% level.

c) We tested for heteroskedasticity (heterogeneity of variance) using White's heteroskedasticity test. If the p-values in this column fall below .050, homoskedasticity is rejected at the 5% level.

d) If homoskedasticity is rejected (see penultimate column and note c), these are robust White-adjusted standard errors.

e) Ramsey RESET test is a general misspecification test for linear regression models. If the p-values in this column fall below .050, the relation between the two bone measures may not be linear.

Table 3b Results of linear regression analysis on bones in Roman stature database (women)

	constant		slope		model			
bones[a]	estimate[b]	S.E.[d]	estimate[b]	S.E.[d]	n	adj. R^2	hetero-scedasticity[c]	Ramsey RESET test[e]
fem1 and tib1	97.693[e]	6.540	.946	.019	1096	.723	.007	.291
fem1 and tib1a	98.630	26.067	.929	.076	38	.802	.868	.319
fem1 and tib1b	77.345	12.611	1.011	.038	385	.730	.001	.991
fem1 and fib1	74.228	11.812	1.037	.036	308	.732	.062	.309
fem1 and hum1	52.753	6.868	1.221	.023	1076	.724	.132	.382
fem1 and hum2	35.478	12.075	1.295	.041	382	.726	.162	.031
fem1 and rad1	145.413	9.769	1.223	.044	915	.541	.000	.000
fem1 and rad1b	89.352	23.322	1.495	.104	122	.631	.513	.999
fem1 and rad2	110.966	23.456	1.460	.109	136	.567	.434	.772
fem1 and uln1	129.980	16.150	1.195	.068	597	.555	.000	.000
fem1 and uln2	184.764	45.439	1.095	.211	123	.411	.000	.000
fem2 and tib1	85.958	9.574	.971	.028	553	.742	.000	.199
fem2 and tib1a	93.216	27.438	.933	.079	36	.796	.945	.328
fem2 and tib1b	86.106	12.120	.973	.036	357	.748	.000	.539
fem2 and fib1	53.214	15.857	1.087	.048	193	.737	.045	.913
fem2 and hum1	52.263	9.828	1.216	.033	510	.730	.186	.116
fem2 and rad1	121.225	12.246	1.321	.056	437	.586	.039	.715
fem2 and rad1b	81.410	26.974	1.511	.120	86	.651	.057	.987
fem2 and uln1	119.582	26.391	1.219	.110	330	.547	.000	.000

Notes:

a) Bone measures are abbreviated in the following way: fem = femur, tib = tibia, fib = fibula, hum = humerus, rad = radius, uln = ulna; Numbers refer to bone measure numbers in Martin (1928).

b) All parameter estimates are significant at the 1% level.

c) We tested for heteroskedasticity (heterogeneity of variance) using White's heteroskedasticity test. If the p-values in this column fall below .050, homoskedasticity is rejected at the 5% level.

d) If homoskedasticity is rejected (see penultimate column and note c), these are robust White-adjusted standard errors.

e) Ramsey RESET test is a general misspecification test for linear regression models. If the p-values in this column fall below .050, the relation between the two bone measures may not be linear.

Table 4a Long bone length proportions in Roman stature database compared to those in popular stature reconstruction methods (men)

bone measures[a]	constant: 95% confidence interval[b]	constant	stature reconstruction method[c]	slope	slope: 95% confidence interval[b]
fem1 and tib1	105.636 to 130.018	75.110	D & H (w)	1.029	0.880 to 0.946
		67.365	D & H (b)	1.029	
		56.296	D & H (g)	1.069	
		181.053	E & al.[d]	0.695	
		-12.716	P[d]	1.264	
		81.500	T	1.000	
		- 4.416	T & G 1952 (w)[f]	1.059	
		- 3.991	T & G 1952 (b)[f]	1.038	
		70.690	T & G 1958 (w)	1.043	
		62.571	T & G 1958 (b)	1.043	
fem1 and tib1a	45.763 to 151.496	174.453	E & al.[e]	0.695	0.776 to 1.082
		-24.849	P[e]	1.264	
fem1 and tib1b	56.493 to 98.197	7.781	Br.	1.209	0.949 to 1.072
		181.053	E & al.[d]	0.695	
		-12.716	P[d]	1.264	
fem1 and fib1	90.604 to 146.378	180.737	E & al.	0.688	0.853 to 1.009
		-88.476	T	1.191	
		43.571	T & G 1952 (w)	1.126	
		72.512	T & G 1952 (b)	1.038	
		42.974	T & G 1958 (w)	1.121	
		37.381	T & G 1958 (b)	1.114	
fem1 and hum1	59.509 to 86.969	101.869	D & H (w)	1.073	1.121 to 1.207
		-19.100	D & H (b)	1.460	
		20.022	D & H (g)	1.327	
		122.316	E & al.	0.990	
		-5.673	P	1.539	
		-89.367	T	1.333	
		37.983	T & G 1952 (w)	1.294	
		-39.100	T & G 1952 (b)	1.545	
		54.181	T & G 1958 (w)	1.246	
		15.524	T & G 1958 (b)	1.371	
fem1 and hum2	34.977 to 84.797	**67.477**	Br	1.651	1.148 to 1.304
fem1 and rad1	106.151 to 138.229	57.706	D & H (w)	1.630	1.276 to 1.406
		56.618	D & H (b)	1.591	
		50.563	D & H (g)	.631	
		61.298	E & al.	1.599	
		27.205	P	1.740	
		73.950	T & G 1952 (w)	1.588	
		53.128	T & G 1952 (b)	1.621	
		59.871	T & G 1958 (w)	1.634	
		62.905	T & G 1958 (b)	1.581	
fem1 and rad1b	24.572 to 107.940	16.900	Br	1.804	1.424 to 1.761

fem1 and rad2	49.433 to 127.826	-82.252	T	**1.619**	1.407 to 1.740
fem1 and uln1	85.360 to 125.356	-20.983	E & al.	1.719	1.228 to 1.377
		53.109	T & G 1952 (w)	1.555	
		42.370	T & G 1952 (b)	1.545	
		43.190	T & G 1958 (w)	1.621	
		50.238	T & G 1958 (b)	1.524	
fem1 and uln2	89.459 to 184.459	-80.700	T	**1.524**	1.149 to 1.546
fem2 and tib1	94.344 to 127.170	180.823	E & al.[d]	0.695	0.878 to 0.968
		-13.035	P[d,e]	1.264	
fem2 and tib1a	44.607 to 116.859	172.153	E & al.[e]	0.695	0.893 to 1.089
		-25.169	P[e]	1.264	
fem2 and tib1b	93.652 to 135.725	180.823	E & al.[d,e]	0.695	0.859 to 0.974
		56.731	O & al.[e]	1.071	
		-13.035	P[d,e]	1.264	
fem2 and fib1	89.497 to 157.322	183.037	E & al.[e]	0.688	0.814 to 1.001
		52.186	O & al.	1.109	
fem2 and hum1	58.055 to 95.341	120.016	E & al.[e]	0.990	1.091 to 1.205
		24.213	O & al.	1.318	
		-53.616	P[e]	1.539	
fem2 and rad1	103.489 to 140.935	58.998	E & al.[e]	1.579	1.260 to 1.412
		26.89	P[e]	1.740	
fem2 and rad1b	44.592 to 152.271	40.493	O & al.	1.726	1.226 to 1.661
fem2 and uln1	77.125 to 127.410	-23.283	E & al.[e]	1.714	1.211 to 1.397
		32.990	O & al.	1.636	

Notes:

a) Bone measures are abbreviated in the following way: fem = femur, tib = tibia, fib = fibula, hum = humerus, rad = radius, uln = ulna; Numbers refer to bone measure numbers in Martin (1928). All measures are (converted) in(to) mm. Most stature reconstruction formulae are based on either the right or the left bone, but recommend taking the average of both sides for stature reconstruction. Thus, if both left and right bone measures are available, we have taken the average of the two. If a stature reconstruction method provides correctives for the use of left vs. right bones, we have adjusted the long bone measure accordingly. For the Olivier (1978) men, the formulae for the left bones have been chosen, in analogy of the Olivier (1978) formulae for women.

b) Confidence intervals are based on OLS regression analysis of Roman stature database. If homoskedasticity is rejected (see table 3), they are computed using robust White-adjusted standard errors. Values that fall within the 95% confidence interval are in bold type.

c) Stature reconstruction methods are abbreviated in the following way: Br = Breitinger (1937), D & H = Dupertuis & Hadden (1951), E & al. = Eliakis & al. (1966), O & al. = Olivier & al. (1978), P = Pearson (1899), T = Telkkä (1950), T & G = Trotter & Gleser (1952) and (1958). Further, (b) stands for ‘blacks’, (w) stands for ‘whites’, and (g) for general formulae.

d) This stature reconstruction method does not differentiate between tibia measurement nr. 1 and tibia measurement nr. 1b. Long bone length proportions therefore are compared to both tibia nr. 1 and tibia nr. 1b figures from the Roman stature database.

e) This stature reconstruction method does not recommend using one or both of these bone measures. However, as it provides a rule of thumb to convert these measures into the recommended bone measures, long bone length proportions can be calculated still.

f) Jantz & al. (1995) have pointed out that Trotter made a mistake measuring the tibia for Trotter & Gleser (1952), erroneously excluding the malleolus. Before application of the 1952 formulae, 11mm should therefore be subtracted from the tibia nr. 1 measure. In calculating the long bone proportion figures for Trotter & Gleser (1952), we have taken this corrective into account. As the formulae for the tibia in Trotter & Gleser (1958) are based on measures both in- and excluding the malleolus, they are unreliable. However, as they were widely used in the past (and as they continue to be used by some), we have included them here anyway.

Table 4b Long bone length proportions in Roman stature database compared to those in popular stature reconstruction methods (women)

bone measures[a]	constant: 95% confidence interval[b]	constant	stature reconstruction method[c]	slope	slope: 95% confidence interval[b]
fem1 and tib1	84.875 to 110.511	38.233	D & H (w)	1.135	0.909 to 0.983
		73.466	D & H (b)	1.009	
		48.166	D & H (g)	1.093	
		-5.135	E & al.[d]	1.232	
		10.317	P[d]	1.209	
		-76.739	T	1.056	
		- 9.907	T & G 1952 (w)[f]	1.174	
		- 6.167	T & G 1952 (b)[f]	1.075	
fem1 and tib1a	45.763 to 151.496	-15.851	E & al.[e]	1.232	0.776 to 1.082
		10.317	P[e]	1.209	
fem1 and tib1b	52.627 to 102.063	-82.102	Ba	1.329	0.937 to 1.085
		-5.135	E & al.[d]	1.232	
		10.317	P[d]	1.209	
fem1 and fib1	50.986 to 197.470	**76.023**	E & al.[g]	**0.992**	0.967 to 1.108
		-83.583	T	1.278	
		22.308	T & G 1952 (w)	1.186	
		84.860	T & G 1952 (b)	**1.092**	
fem1 and hum1	39.276 to 66.230	-63.290	Ba	1.615	1.175 to 1.266
		24.143	D & H (w)	1.485	
		64.858	D & H (b)[g]	**1.215**	
		15.386	D & H (g)	1.357	
		122.819	E & al.	0.961	
		-3.578	P	1.416	
		-87.850	T	1.500	
		15.668	T & G 1952 (w)	1.360	
		21.535	T & G 1952 (b)	1.351	
fem1 and hum2	11.735 to 59.221	-55.217	Ba	1.615	1.215 to 1.376
fem1 and rad1	126.266 to 164.560	25.368	D & H (w)	1.834	1.137 to 1.309
		85.225	D & H (b)	1.506	
		52.1780	D & H (g)	1.673	
		-136.795	E & al.	2.490	
		44.568	P	1.719	
		3.360	T & G 1952 (w)	1.919	
		52.763	T & G 1952 (b)	1.610	
fem1 and rad1b	43.177 to 135.527	**77.685**	Ba[g]	**1.466**	1.290 to 1.701
fem1 and rad2	64.556 to 157.375	-77.622	T	1.466	1.243 to 1.676
fem1 and uln1	98.326 to 161.634	-10.232	E & al.	1.772	1.062 to 1.328
		14.818	T & G 1952 (w)	1.729	
		68.509	T & G 1952 (b)	1.452	

fem1 and uln2	95.704 to 273.824	-80.850	T	1.833	0.681 to 1.709
fem2 and tib1	67.193 to 104.723	-8.435 9.987	E & al.[d,e] P[d,e]	1.232 1.209	0.916 to 1.026
fem2 and tib1a	37.545 to 148.978	-19.151 -0.534	E & al.[e] P[e]	1.232 1.209	0.772 to 1.094
fem2 and tib1b	62.351 to 109.861	-8.435 48.664 9.987	E & al.[d,e] O & al. P[d,e]	1.232 1.097 1.209	0.902 to 1.044
fem2 and fib1	22.134 to 84.294	**79.323** **52.186**	E & al.[e,g] O & al.[g]	0.992 **1.109**	0.993 to 1.181
fem2 and hum1	32.954 to 71.572	126.119 -37.643 -3.908	E & al.[e] O & al. P[e]	0.961 1.473 1.416	1.151 to 1.280
fem2 and rad1	97.223 to 145.227	**138.095** 44.238	E & al.[e] P[e]	2.490 1.719	1.211 to 1.431
fem2 and rad1b	27.769 to 135.051	0.477	O & al.	1.972	1.273 to 1.749
fem2 and uln1	67.856 to 171.308	-13.532 -32.917	E & al.[e] O & al.	1.772 1.953	1.003 to 1.435

Notes:

a) Bone measures are abbreviated in the following way: fem = femur, tib = tibia, fib = fibula, hum = humerus, rad = radius, uln = ulna; Numbers refer to bone measure numbers in Martin (1928). All measures are (converted) in(to) mm. Most stature reconstruction formulae are based on either the right or the left bone, but recommend taking the average of both sides for stature reconstruction. Thus, if both left and right bone measures are available, we have taken the average of the two. If a stature reconstruction method provides correctives for the use of left vs. right bones, we have adjusted the long bone measure accordingly. For the Olivier (1978) men, the formulae for the left bones have been chosen, in analogy of the Olivier (1978) formulae for women.

b) Confidence intervals are based on OLS regression analysis of Roman stature database. If homoskedasticity is rejected (see table 3), they are computed using robust White-adjusted standard errors. Values that fall within the 95% confidence interval are in bold type.

c) Stature reconstruction methods are abbreviated in the following way: Br = Breitinger (1937), D & H = Dupertuis & Hadden (1951), E & al. = Eliakis & al. (1966), O & al. = Olivier & al. (1978), P = Pearson (1899), T = Telkkä (1950), T & G = Trotter & Gleser (1952) and (1958). Further, (b) stands for 'blacks', (w) stands for 'whites', and (g) for general formulae.

d) This stature reconstruction method does not differentiate between tibia measurement nr. 1 and tibia measurement nr. 1b. Long bone length proportions therefore are compared to both tibia nr. 1 and tibia nr. 1b figures from the Roman stature database.

e) This stature reconstruction method does not recommend using one or both of these bone measures. However, as it provides a rule of thumb to convert these measures into the recommended bone measures, long bone length proportions can be calculated still.

f) Jantz and al. (1995) have pointed out that Trotter made a mistake measuring the tibia for Trotter & Gleser (1952), erroneously excluding the malleolus. Before application of the 1952 formulae, 11mm should therefore be subtracted from the tibia nr. 1 measure. In calculating the long bone proportion figures for Trotter & Gleser (1952), we have taken this corrective into account. As the formulae for the tibia in Trotter & Gleser (1958) are based on measures both in- and excluding the malleolus, they are unreliable. However, as they were widely used in the past (and as they continue to be used by some), we have included them here anyway.

g) As both slope and constant (almost) fall within the 95% confidence interval, both parameters have been tested together using the Wald test. In all cases, they were significantly different from the values for the Roman stature database (p = .000).

Table 5a Robustness test: Results of linear regression analysis on bones in Roman stature database (men)

	constant		slope		model			
								Ramsey
						adj.	hetero-	RESET
bones[a]	estimate[b]	S.E.[d]	estimate[b]	S.E.[d]	n	R^2	skedasticity[c]	test[e]
fem1 and tib1[d]	1.576	6.170	.807	.014	1349	.737	.005	.073
fem1 and tib1a	19.843	17.746	.775	.040	96	.801	.725	.332
fem1 and tib1b	6.379	10.254	.791	.023	432	.739	.145	.984
fem1 and fib1	19.148	12.100	.751	.027	343	.698	.618	.558
fem1 and hum1	60.175	5.444	.587	.012	1398	.683	.000	.192
fem1 and hum2	68.936	7.241	.556	.016	571	.681	.425	.507
fem1 and rad1	32.743	5.590	.472	.012	1127	.633	.000	.044
fem1 and rad1b	45.701	10.794	.438	.023	153	.695	.562	.656
fem1 and rad2	38.979	10.522	.427	.023	171	.670	.551	.760
fem1 and uln1	56.059	6.182	.466	.014	762	.606	.127	.648
fem1 and uln2	57.510	13.530	.395	.029	160	.529	.852	.883
fem2 and tib1	10.194	8.759	.795	.020	751	.733	.003	.016
fem2 and tib1a	15.183	17.634	.792	.040	112	.783	.882	.420
fem2 and tib1b	10.031	11.348	.790	.025	375	.723	.177	.904
fem2 and fib1	51.480	16.254	.687	.036	223	.622	.066	.004
fem2 and hum1	57.993	7.718	.594	.017	727	.682	.003	.175
fem2 and rad1	22.032	6.495	.497	.014	607	.663	.873	.713
fem2 and rad1b	29.224	16.514	.479	.036	80	.687	.473	.455
fem2 and uln1	55.708	7.779	.471	.017	476	.613	.628	.574

Notes:

a) Bone measures are abbreviated in the following way: fem = femur, tib = tibia, fib = fibula, hum = humerus, rad = radius, uln = ulna; Numbers refer to bone measure numbers in Martin (1928).

b) All parameter estimates are significant at the 1% level.

c) We tested for heteroskedasticity (heterogeneity of variance) using White's heteroskedasticity test. If the p-values in this column fall below .050, homoskedasticity is rejected at the 5% level.

d) If homoskedasticity is rejected (see penultimate column and note c), these are robust White-adjusted standard errors.

e) Ramsey RESET test is a general misspecification test for linear regression models. If the p-values in this column fall below .050, the relation between the two bone measures may not be linear.

Table 5b Robustness test: Results of linear regression analysis on bones in Roman stature database (women)

	constant		slope		model			
bones[a]	estimate[b]	S.E.[d]	estimate[b]	S.E.[d]	n	adj. R^2	hetero-skedasticity[c]	Ramsey RESET test[e]
fem1 and tib1[d]	18.992	5.985	.765	.014	1096	.723	.059	.239
fem1 and tib1a	-19.487	29.641	.870	.071	38	.802	.516	.532
fem1 and tib1b	34.938	9.407	.723	.022	385	.730	.646	.990
fem1 and fib1	35.635	10.153	.706	.024	308	.732	.005	.174
fem1 and hum1	51.008	4.665	.593	.011	1076	.724	.528	.711
fem1 and hum2	60.962	7.403	.561	.018	382	.726	.873	.502
fem1 and rad1	37.176	5.619	.443	.013	915	.541	.067	.002
fem1 and rad1b	44.283	12.519	.424	.029	122	.631	.167	.717
fem1 and rad2	48.564	12.427	.391	.029	136	.567	.000	.000
fem1 and uln1	46.486	7.122	.465	.017	597	.555	.300	.620
fem1 and uln2	57.198	17.359	.380	.041	123	.411	.629	.197
fem2 and tib1	21.956	8.008	.765	.019	553	.742	.526	.502
fem2 and tib1a	-11.933	30.416	.860	.073	36	.796	.383	.691
fem2 and tib1b	18.601	9.832	.770	.024	357	.748	.371	.399
fem2 and fib1	50.602	12.128	.680	.029	193	.737	.016	.039
fem2 and hum1	49.295	6.757	.601	.016	510	.730	.862	.700
fem2 and rad1	37.887	7.426	.444	.018	437	.586	.003	.001
fem2 and rad1b	42.389	14.487	.433	.034	86	.651	.958	.077
fem2 and uln1	55.996	9.389	.450	.023	330	.547	.539	.034

Notes:

a) Bone measures are abbreviated in the following way: fem = femur, tib = tibia, fib = fibula, hum = humerus, rad = radius, uln = ulna; Numbers refer to bone measure numbers in Martin (1928).

b) All parameter estimates are significant at the 1% level.

c) We tested for heteroskedasticity (heterogeneity of variance) using White's heteroskedasticity test. If the p-values in this column fall below .050, homoskedasticity is rejected at the 5% level.

d) If homoskedasticity is rejected (see penultimate column and note c), these are robust White-adjusted standard errors.

e) Ramsey RESET test is a general misspecification test for linear regression models. If the p-values in this column fall below .050, the relation between the two bone measures may not be linear.

Table 6a Robustness test: long bone length proportions in Roman stature database compared to those in popular stature reconstruction methods (men)

bone measures[a]	constant: 95% confidence interval[b]	constant	stature reconstruction method[c]	slope	slope: 95% confidence interval[b]
tib1 and fem1	-10.517 to 13.669	-72,993	D & H (w)	0,972	.780 to .834
		-65,466	D & H (b)	0,972	
		-52,662	D & H (g)	0,935	
		-260,508	E & al.[d]	1,439	
		10,060	P[d,g]	**0,791**	
		-81,500	T	1.000	
		4,170	T & G 1952 (w)[f]	0,944	
		3,845	T & G 1952 (b)[f]	0,963	
		-67,776	T & G 1958 (w)	0,959	
		-59,991	T & G 1958 (b)	0,959	
tib1a and fem1	-15.393 to 55.079	-251,012	E & al.[e]	1,439	.697 to .854
		19,659	P[e,g]	**0,791**	
tib1b and fem1	-13.775 to 26.533	**-6,436**	Br[f,g]	**0,827**	.747 to .836
		-260,508	E & al.[d]	1,439	
		10,060	P[d,g]	**0,791**	
fib1 and fem1	-4.652 to 42.947	-262,699	E & al.	1,453	.698 to .803
		74,287	T	0,840	
		-38,695	T & G 1952 (w)	0,888	
		-69,857	T & G 1952 (b)	0,963	
		-38,335	T & G 1958 (w)	0,892	
		-33,556	T & G 1958 (b)	0,898	
hum1 and fem1	49.513 to 70.837	-94,938	D & H (w)	0,932	.565 to .611
		13,082	D & H (b)	0,685	
		-15,088	D & H (g)	0,754	
		-123,552	E & al.	1,010	
		3,686	P	0,650	
		67,042	T	0,750	
		-29,353	T & G 1952 (w)	0,773	
		25,307	T & G 1952 (b)	0,647	
		-43,484	T & G 1958 (w)	0,803	
		-11,323	T & G 1958 (b)	0,729	
hum2 and fem1	54.714 to 83.158	-40,87	Br	0,606	.525 to .588
rad1 and fem1	21.787 to 43.699	-35,402	D & H (w)	0,613	.448 to .496
		-35,586	D & H (b)	0,629	
		-80,132	D & H (g)	1,585	
		-38,335	E & al.	0,625	
		-15,635	P	0,575	
		-46,568	T & G 1952 (w)	0,630	
		-32,775	T & G 1952 (b)	0,617	
		-36,641	T & G 1958 (w)	0,612	
		-39,788	T & G 1958 (b)	0,633	
rad1b and fem1	24.374 to 67.029	-9,368	Br	0,554	.391 to .484

rad2 and fem1	18.209 to 59.750	50,804	T	0,618	.382 to .473
uln1 and fem1	43.923 to 68.196	12,207 -34,154 -27,424 -26,644 -32,965	E & al. T & G 1952 (w) T & G 1952 (b) T & G 1958 (w) T & G 1958 (b)	0,582 0,643 0,647 0,617 0,656	.439 to .493
uln2 and fem1	30.786 to 84.233	**52,953**	T	0,656	.337 to .453
tib1 and fem2	-6.974 to 27.362	-260,177 **10,313**	E & al.[d] P[d,e,f,g]	1,439 **0,791**	.756 to .834
tib1a and fem2	-19.763 to 50.130	-247,702 **19,912**	E & al.[e] P[e,f,g]	1,439 **0,791**	.714 to .871
tib1b and fem2	-12.283 to 32.345	-260,177 -52,97 **10,313**	E & al.[d,e] O & al.[e] P[d,e,g]	1,439 0,934 **0,791**	.740 to.840
fib1 and fem2	19.446 to 83.513	-266,042 -47,057	E & al.[e] O & al.	1,453 0,902	.616 to.758
hum1 and fem2	42.866 to 73.120	-121,228 -18,371 34,838	E & al.[e] O & al. P[e]	1,010 0,759 0,650	.561 to .627
rad1 and fem2	9.276 to 34.788	-37,364 -1,545	E & al.[e] P[e]	0,633 0,575	.469 to .525
rad1b and fem2	-3.653 to 62.101	-23,461	O & al.	0,579	.406 to .551
uln1 and fem2	40.423 to 70.993	13,584 -20,165	E & al.[e] O & al.	0,583 0,611	.437 to.504

Notes:

a) Bone measures are abbreviated in the following way: fem = femur, tib = tibia, fib = fibula, hum = humerus, rad = radius, uln = ulna; Numbers refer to bone measure numbers in Martin (1928). All measures are (converted) in(to) mm. Most stature reconstruction formulae are based on either the right or the left bone, but recommend taking the average of both sides for stature reconstruction. Thus, if both left and right bone measures are available, we have taken the average of the two. If a stature reconstruction method provides correctives for the use of left vs. right bones, we have adjusted the long bone measure accordingly. For the Olivier (1978) men, the formulae for the left bones have been chosen, in analogy of the Olivier (1978) formulae for women.

b) Confidence intervals are based on OLS regression analysis of Roman stature database. If homoskedasticity is rejected (see table 3), they are computed using robust White-adjusted standard errors. Values that fall within the 95% confidence interval are in bold type.

c) Stature reconstruction methods are abbreviated in the following way: Br = Breitinger (1937), D & H = Dupertuis & Hadden (1951), E & al. = Eliakis & al. (1966), O & al. = Olivier & al. (1978), P = Pearson (1899), T = Telkkä (1950), T & G = Trotter & Gleser (1952) and (1958). Further, (b) stands for 'blacks', (w) stands for 'whites', and (g) for general formulae.

d) This stature reconstruction method does not differentiate between tibia measurement nr. 1 and tibia measurement nr. 1b. Long bone length proportions therefore are compared to both tibia nr. 1 and tibia nr. 1b figures from the Roman stature database.

e) This stature reconstruction method does not recommend using one or both of these bone measures. However, as it provides a rule of thumb to convert these measures into the recommended bone measures, long bone length proportions can be calculated still.

f) Jantz and al. (1995) have pointed out that Trotter made a mistake measuring the tibia for Trotter & Gleser (1952), erroneously excluding the malleolus. Before application of the 1952 formulae, 11mm should therefore be subtracted from the tibia nr. 1 measure. In calculating the long bone proportion figures for Trotter & Gleser (1952), we have taken this corrective into account. As the formulae for the tibia in Trotter & Gleser (1958) are based on measures both in- and excluding the malleolus, they are unreliable. However, as they were widely used in the past (and as they continue to be used by some), we have included them here anyway.

g) As both slope and constant (almost) fall within the 95% confidence interval, both parameters have been tested together using the Wald test. In all cases, they were significantly different from the values for the Roman stature database (p = .000), except tibia measure 1b and femur measure 2 in Pearson (p = .407).

Table 6b Robustness test: Long bone length proportions in Roman stature database compared to those in popular stature reconstruction methods (women)

bone measures[a]	constant 95% confidence interval[b]		stature reconstruction method[c]	slope	95% confidence interval[b]
tib1 and fem1	7.250 to 30.735	-33,685	D & H (w)	0,881	.737 to .739
		-72,811	D & H (b)	0,991	
		-44,068	D & H (g)	0,915	
		4,168	E & al.[d]	0,812	
		-8,533	P[d]	0,827	
		72,67	T	0,947	
		8,439	T & G 1952 (w)[f]	0,852	
		5,737	T & G 1952 (b)[f]	0,930	
tib1a and fem1	-79.602 to 40.628	**12,866**	E & al.[e,g]	**0,812**	.726 to 1.013
		-8,533	P[e,g]	**0,827**	
tib1b and fem1	16.443 to 53.434	61,777	Ba	**0,752**	.679 to .767
		4,168	E & al.[d]	0,812	
		-8,533	P[d]	0,827	
fib1 and fem1	14.916 to 56.354	-76,636	E & al.	1,008	.655 to .757
		65,401	T	0,782	
		-18,809	T & G 1952 (w)	0,843	
		-77,711	T & G 1952 (b)	0,916	
hum1 and fem1	41.854 to 60.162	39,189	Ba	0,619	.571 to .615
		-16,258	D & H (w)	0,673	
		-53,381	D & H (b)[g]	0,823	
		-11,338	D & H (g)	0,737	
		-127,803	E & al.	1,041	
		2,527	P	0,706	
		58,567	T	0,667	
		-11,521	T & G 1952 (w)	0,735	
		-15,94	T & G 1952 (b)	0,740	
hum2 and fem1	46.407 to 75.517	34,19	Ba	0,619	.526 to .596
rad1 and fem1	26.148 to 84.204	-13,832	D & H (w)	0,545	.416 to .469
		-56,59	D & H (b)	0,664	
		-311,883	D & H (g)	0,598	
		54,938	E & al.	0,402	
		-25,927	P	0,582	
		-1,751	T & G 1952 (w)	0,521	
		-32,772	T & G 1952 (b)	0,621	
rad1b and fem1	19.746 to 68.820	-52,991	Ba[g]	0,682	.367 to .481
rad2 and fem1	14.619 to 82.509	**52,948**	T	0,682	.311 to .471
uln1 and fem1	32.498 to 60.473	5,774	E & al.	0,564	.431 to .498
		-8,57	T & G 1952 (w)	0,578	
		-47,183	T & G 1952 (b)	0,689	
uln2 and fem1	22.831 to 91.565	**44,108**	T	0,546	.299 to .461

tib1 and fem2	6.227 to 37.686	**6,847** -8,261	E & al.[d,e] P[d,e]	0,812 **0,827**	.727 to .802
tib1a and fem2	-73.746 to 49.881	**15,545** **0,442**	E & al.[e,g] P[e,g]	**0,812** **0,827**	.711 to 1.009
tib1b and fem2	-0.734 to 37.937	**6,847** -44,361 -8,261	E & al.[d,e,g] O & al. P[d,e]	**0,812** 0,912 0,827	.723 to .816
fib1 and fem2	26.261 to 75.943	-79,963 -47,057	E & al.[e] O & al.	1,008 0,902	.619 to .741
hum1 and fem2	36.021 to 62.569	-131,237 25,555 2,760	E & al.[e] O & al. P[e]	1,041 0,679 0,706	.569 to .633
rad1 and fem2	20.796 to 54.978	-55,460 -25,735	E & al.[e] P[e]	0,402 0,582	.403 to .485
rad1b and fem2	13.580 to 71.197	-0,242	O & al.	0,507	.365 to .502
uln1 and fem2	37.526 to 74.466	7,637 16.856	E & al.[e] O & al.	0,564 0,512	.405 to.494

Notes:

a) Bone measures are abbreviated in the following way: fem = femur, tib = tibia, fib = fibula, hum = humerus, rad = radius, uln = ulna; Numbers refer to bone measure numbers in Martin (1928). All measures are (converted) in(to) mm. Most stature reconstruction formulae are based on either the right or the left bone, but recommend taking the average of both sides for stature reconstruction. Thus, if both left and right bone measures are available, we have taken the average of the two. If a stature reconstruction method provides correctives for the use of left vs. right bones, we have adjusted the long bone measure accordingly. For the Olivier (1978) men, the formulae for the left bones have been chosen, in analogy of the Olivier (1978) formulae for women.

b) Confidence intervals are based on OLS regression analysis of Roman stature database. If homoskedasticity is rejected (see table 3), they are computed using robust White-adjusted standard errors. Values that fall within the 95% confidence interval are in bold type.

c) Stature reconstruction methods are abbreviated in the following way: Br = Breitinger (1937), D & H = Dupertuis & Hadden (1951), E & al. = Eliakis & al. (1966), O & al. = Olivier & al. (1978), P = Pearson (1899), T = Telkkä (1950), T & G = Trotter & Gleser (1952) and (1958). Further, (b) stands for 'blacks', (w) stands for 'whites', and (g) for general formulae.

d) This stature reconstruction method does not differentiate between tibia measurement nr. 1 and tibia measurement nr. 1b. Long bone length proportions therefore are compared to both tibia nr. 1 and tibia nr. 1b figures from the Roman stature database.

e) This stature reconstruction method does not recommend using one or both of these bone measures. However, as it provides a rule of thumb to convert these measures into the recommended bone measures, long bone length proportions can be calculated still.

f) Jantz and al. (1995) have pointed out that Trotter made a mistake measuring the tibia for Trotter & Gleser (1952), erroneously excluding the malleolus. Before application of the 1952 formulae, 11mm should therefore be subtracted from the tibia nr. 1 measure. In calculating the long bone proportion figures for Trotter & Gleser (1952), we have taken this corrective into account. As the formulae for the tibia in Trotter & Gleser (1958) are based on measures both in- and excluding the malleolus, they are unreliable. However, as they were widely used in the past (and as they continue to be used by some), we have included them here anyway.

g) As both slope and constant (almost) fall within the 95% confidence interval, both parameters have been tested together using the Wald test. In all cases, they were significantly different from the values for the Roman stature database (p = .000), except tibia measure 1a and femur measure 2 in Pearson (p = .618).

List of research reports

12001-HRM&OB: Veltrop, D.B., C.L.M. Hermes, T.J.B.M. Postma and J. de Haan, A Tale of Two Factions: Exploring the Relationship between Factional Faultlines and Conflict Management in Pension Fund Boards

12002-EEF: Angelini, V. and J.O. Mierau, Social and Economic Aspects of Childhood Health: Evidence from Western-Europe

12003-Other: Valkenhoef, G.H.M. van, T. Tervonen, E.O. de Brock and H. Hillege, Clinical trials information in drug development and regulation: existing systems and standards

12004-EEF: Toolsema, L.A. and M.A. Allers, Welfare financing: Grant allocation and efficiency

12005-EEF: Boonman, T.M., J.P.A.M. Jacobs and G.H. Kuper, The Global Financial Crisis and currency crises in Latin America

12006-EEF: Kuper, G.H. and E. Sterken, Participation and Performance at the London 2012 Olympics

12007-Other: Zhao, J., G.H.M. van Valkenhoef, E.O. de Brock and H. Hillege, ADDIS: an automated way to do network meta-analysis

12008-GEM: Hoorn, A.A.J. van, Individualism and the cultural roots of management practices

12009-EEF: Dungey, M., J.P.A.M. Jacobs, J. Tian and S. van Norden, On trend-cycle decomposition and data revision

12010-EEF: Jong-A-Pin, R., J-E. Sturm and J. de Haan, Using real-time data to test for political budget cycles

12011-EEF: Samarina, A., Monetary targeting and financial system characteristics: An empirical analysis

12012-EEF: Alessie, R., V. Angelini and P. van Santen, Pension wealth and household savings in Europe: Evidence from SHARELIFE

13001-EEF: Kuper, G.H. and M. Mulder, Cross-border infrastructure constraints, regulatory measures and economic integration of the Dutch – German gas market

13002-EEF: Klein Goldewijk, G.M. and J.P.A.M. Jacobs, The relation between stature and long bone length in the Roman Empire www.rug.nl/feb